

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-320394

(43)Date of publication of application : 04.12.1998

(51)Int.Cl.

G06F 17/28

(21)Application number : 09-128673

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 19.05.1997

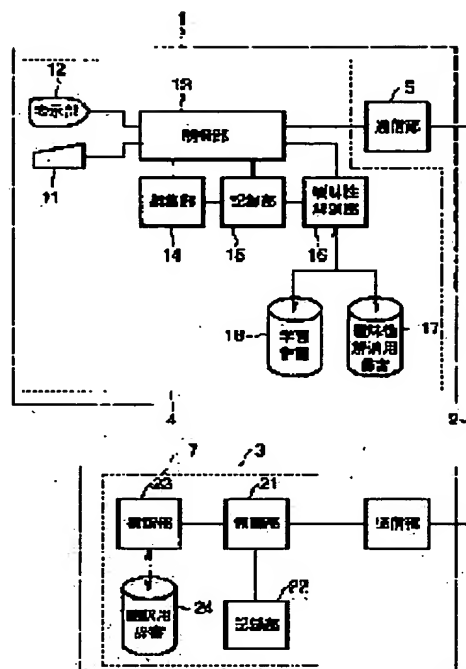
(72)Inventor : NAKAMURA MARIKO
ITO ETSUO

(54) NATURAL LANGUAGE PROCESSOR

(57)Abstract:

PROBLEM TO BE SOLVED: To reduce the load of a server terminal or a network, to efficiently use a server client, and to operate a natural language processing.

SOLUTION: At least one processing request terminal 1 makes a natural language processing request. A natural language processing is operated in response to the natural language processing request of this processing request terminal 1 at least one natural language processing terminal 3, and in the case of translation with vagueness whose translated result can not be uniquely decided generated during this natural language processing, identification data are added to the translation with ambiguity without canceling ambiguity, and it is transmitted with the other translated results as it is. Moreover, when the processing requesting terminal 1 receives the translation with ambiguity from the natural language processing terminal 3, an ambiguity canceling means 16 generates another candidate by referring to an ambiguity cancellation dictionary 17, selects one of the other candidates, and preserves it in its own learning dictionary 18 without transmitting it to the natural language processing terminal 3.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

This Page Blank (uspto)

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

This Page Blank (uspto)

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平10-320394

(43)公開日 平成10年(1998)12月4日

(51)Int.Cl.⁶

G 0 6 F 17/28

識別記号

F I

G 0 6 F 15/38

Q

審査請求 未請求 請求項の数6 O L (全 8 頁)

(21)出願番号 特願平9-128673

(22)出願日 平成9年(1997)5月19日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 中村 真理子

東京都府中市東芝町1番地 株式会社東芝
府中工場内

(72)発明者 伊藤 悦雄

東京都府中市東芝町1番地 株式会社東芝
府中工場内

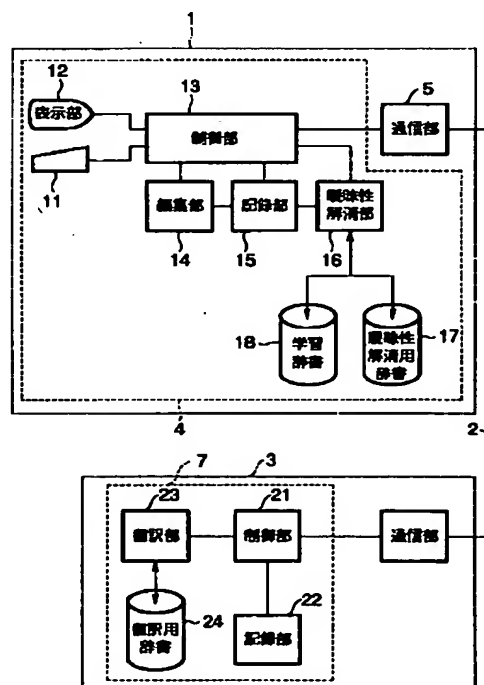
(74)代理人 弁理士 鈴江 武彦 (外6名)

(54)【発明の名称】 自然言語処理装置

(57)【要約】

【課題】 サーバ端末やネットワークの負荷低減化およびサーバ・クライアントを効率よく利用しつつ自然言語処理を行うことにある。

【解決手段】 自然言語処理要求を行う少なくとも1台の処理要求端末1と、この処理要求端末の自然言語処理要求に応じて自然言語処理を行うとともに、この自然言語処理中に生ずる翻訳結果を一意に決定できない曖昧性の存在する訳文に関しては曖昧性を解消せずに曖昧性の存在する訳文に識別データを付けてそのまま他の翻訳結果データと共に送信する少なくとも1台の自然言語処理端末3とを設け、さらに、前記処理要求端末は、自然言語処理端末から曖昧性を含んだ訳文を受けたとき、曖昧性解消用辞書を参照しつつ他候補を生成するとともに、他候補の中から1つを選択したとき、自然言語処理端末に送信せずに自己の学習辞書に保存する曖昧性解消手段16を設けた自然言語処理装置である。



【特許請求の範囲】

【請求項1】 複数の端末がネットワークを介して電子的に接続されてなる自然言語処理装置において、自然言語処理要求を行う前記端末一方となる少なくとも1台の処理要求端末と、この処理要求端末の自然言語処理要求に応じて自然言語処理を行うとともに、この自然言語処理中に生ずる翻訳結果を一意に決定できない曖昧性の存在する訳文に関しては曖昧性を解消せずにそのまま他の翻訳結果データと共に送信する前記端末の他方となる少なくとも1台の自然言語処理端末とを設けたことを特徴とする自然言語処理装置。

【請求項2】 自然言語処理端末は、曖昧性の存在する訳文に、曖昧性の存在しない訳文とは区別できる識別データを設定して送信することを特徴とする請求項1に記載の自然言語処理装置。

【請求項3】 自然言語処理端末は、曖昧性を含む文について、その文中の何れの個所に曖昧性が有るかを示す情報を送信することを特徴とする請求項1に記載の自然言語処理装置。

【請求項4】 処理要求端末は、前記自然言語処理端末から曖昧性を含んだ訳文を受けたとき、曖昧性解消用辞書部の辞書を参照しつつ候補を生成するとともに、他候補の中から1つを選択したとき、前記自然言語処理端末に送信せずに自己の学習辞書に保存する曖昧性解消手段を有することを特徴とする請求項1に記載の自然言語処理装置。

【請求項5】 曖昧性解消手段は、訳文に含まれる曖昧性に対し、ユーザカスタマイズ項目を設定することにより、曖昧性解消処理データのデータ量を低減化することを特徴とする請求項4に記載の自然言語処理装置。

【請求項6】 曖昧性解消手段は、曖昧性解消用データを生成する際、ユーザカスタマイズ項目を設定することにより、ユーザが選択しやすいデータ表示とし、かつ、曖昧性解消用データのデータ量を低減化することを特徴とする請求項4に記載の自然言語処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ネットワーク上の複数の端末が相互に分担分けして自然言語処理を行う自然言語処理装置に関する。

【0002】

【従来の技術】近年、コンピュータネットワークの発達に伴い、同一ネットワーク上に多種類の計算機が多数接続され、また電話網のデジタル化や当該電話網の付随機器の整備・開発等によって通信速度も大きく向上し、計算機相互の間で大量のデータを容易に授受可能になってきた。

【0003】ところで、ネットワーク上に接続される多数の計算機は、全て高い能力をもっているとは限らない

ことから、能力の高い計算機と能力の低い計算機とを接続し、ジョブの分担を行いながらデータの授受を行うサーバ・クライアントモデルが普及してきている。

【0004】このようなサーバ・クライアントモデルでは、機械翻訳、音声認識、要約、かな漢字変換、文書校正支援などの自然言語処理に関し、能力の高いサーバ・マシンは、計算機パワーを必要とする自然言語処理を担当し、一方、能力の低いクライアント・マシンでは、専らサーバ・マシンに自然言語処理要求を行うといった、自然言語処理方法をとっている。

【0005】ちなみに、ネットワークを利用した機械翻訳処理システムには、ネットワーク間交信処理装置（特開平4-127748号公報）、機械翻訳ネットワークシステム（特開平5-128144号公報）、自然言語処理装置（特願平5-348542号）などがある。

【0006】これらシステムの自然言語処理時の曖昧性解消処理としては、前者の2つの出願、つまり「ネットワーク間交信処理装置」、「機械翻訳ネットワークシステム」では、サーバ・マシンが曖昧性の生じた箇所に対して1つの解を決定し、この決定性データのみを許す送信手段により、クライアント・マシンに転送する方法をとっている。

【0007】また、後者の「自然言語処理装置」は、サーバ・マシンが曖昧性を含んだ箇所に対する可能な解の生成を行い、その解全てを非決定性データとして他の翻訳結果データと共にクライアント・マシンに転送する方法をとっている。

【0008】

【発明が解決しようとする課題】従って、以上のようなシステムのうち、前者の2つの出願は、サーバ・マシンから決定性データを転送することから、クライアント・マシンでは、例えば機械翻訳やかな漢字変換等によって他の翻訳候補や次候補などを取得することができない問題がある。一方、後者の出願では、多数のクライアント端末から翻訳要求が出されたとき、サーバ・マシンやネットワークの負荷が増大するといった問題がある。

【0009】本発明は、上記実情に鑑みてなされたもので、サーバ端末やネットワークの負荷を低減化し、効率よく自然言語処理を行う自然言語処理装置を提供することを目的とする。

【0010】

【課題を解決するための手段】請求項1に対応する発明は、上記課題を解決するために、複数の端末がネットワークを介して電子的に接続されてなる自然言語処理装置において、自然言語処理要求を行う前記端末一方となる少なくとも1台の処理要求端末と、この処理要求端末の自然言語処理要求に応じて自然言語処理を行うとともに、この自然言語処理中に生ずる翻訳結果を一意に決定できない曖昧性の存在する訳文に関しては曖昧性を解消

せずにそのまま他の翻訳結果データと共に送信する前記端末の他方となる少なくとも1台の自然言語処理端末とを設けた自然言語処理装置である。

【0011】この請求項1に対応する発明では、処理要求端末から自然言語処理要求を受けると、自然言語処理端末は、自然言語処理を実行するが、この自然言語処理中に翻訳結果を一意に決定できない曖昧性の存在する訳文が生じたとき、曖昧性を解消せずにそのまま他の翻訳結果データと共に送信するので、自然言語処理端末の処理負担が少くなり、またネットワークの占有時間が大幅に低減することにより、多数の処理要求端末の要求処理に十分対処でき、一方、処理要求端末では、十分な時間をかけて曖昧性を解消する処理を行うことができ、各端末を効率よく利用しながら自然言語処理を行うことができる。

【0012】なお、自然言語処理端末は、曖昧性の存在する訳文に、曖昧性の存在しない訳文とは区別できる識別データ例えばマーキング、×マーク、星マークを設定するとか、文中の何れの個所に曖昧性があるかを示す情報例えば何頁何行目とかの情報を送信して送信すれば、曖昧性を含む翻訳結果のデータを処理要求端末に確実に転送でき、また処理要求端末では識別データから容易に曖昧性の訳文を見つけ出して曖昧性の解消処理を行うことができる。

【0013】また、処理要求端末は、自然言語処理端末から曖昧性を含んだ訳文を受けたとき、曖昧性解消用辞書を参照しつつ他候補を生成し、他候補の中から1つを選択したとき、前記自然言語処理端末に送信せずに自己の学習辞書に保存する曖昧性解消手段を設ければ、後の曖昧性解消時に学習結果を有効に利用できる。

【0014】さらに、曖昧性解消手段は、訳文に含まれる曖昧性に対してユーザカスタマイズ項目を設定するとか、或いは曖昧性解消用データを生成する際してユーザカスタマイズ項目を設定することにより、ユーザが選択しやすいデータ表示とし、かつ、曖昧性解消用データのデータ量を低減するようにする。きる。

【0015】

【発明の実施の形態】図1は本発明に係わる自然言語処理装置の一実施の形態である機械翻訳処理装置の概略構成を示す図である。この機械翻訳処理装置は、大きく分けて、自然言語処理要求を行う処理要求端末1と、この処理要求端末1に通信回線となるネットワーク2を介して接続され、処理要求端末1から送られてくる自然言語処理要求に基づいて自然言語処理を行う自然言語処理端末3とからなっている。

【0016】この処理要求端末1は、翻訳用原文を入力表示したり、曖昧性解消処理を行うデータ処理制御部4および自然言語処理端末3との間でデータの授受を行う通信部5を有し、一方、自然言語処理端末3は、処理要求端末1との間でデータの授受を行う通信部6および翻

訳用原文の言語に対する翻訳処理を実行する自然言語処理エンジン7をもっている。

【0017】さらに、図1に示す各部4～7は、図2に示すような構成となっている。なお、処理要求端末1はクライアント端末に相当し、自然言語処理端末3はサーバ端末に相当し、これらはそれぞれ電子計算機が用いられている。

【0018】以下、各部4～7の構成を具体的に説明する。処理要求端末1の中のデータ処理制御部4は、入力部11および表示部12が設けられている。入力部11は、処理要求を行う翻訳用原文の言語や処理要求コマンドを入力したり、処理結果を修正するための修正データの入力、訳文候補の選択を行うものであって、例えばキーボード、マウス、タッチパネル、ペンなどのポインティングデバイスによって構成されている。表示部12は、入力部11から入力されるコマンドや文字列を表示したり、翻訳結果を表示するものであって、例えばCRTディスプレイやビットマップディスプレイなどで構成されるが、ゴーグルタイプやプロジェクタタイプのものであってもよい。

【0019】前記入力部11の入力データは制御部13に送られる。この制御部13は、入力部11および表示部12の他、編集部14、記録部15、曖昧性解消部16および通信部5などと接続され、後記する図3の処理制御やデータの受け渡しを行う機能をもっている。

【0020】前記編集部14は、入力部11から入力されるデータや受信部5で受信された翻訳データを選択するために用いられる。記録部15は、入力部11から入力されるデータ、通信部5で受信された翻訳データ、曖昧性解消部16から得られた曖昧性解消用の選択肢、編集部14によって編集されたデータを記録するために用いられる。この曖昧性解消部16は、通信部5で受信された翻訳データの中に曖昧性が含まれている場合、曖昧性解消用辞書17の辞書を参照しつつ他候補を生成し、表示部12に表示するなどしながら、対話的に1つの訳語を決定する機能をもっている。ここで、曖昧性解消用辞書17に保存されている辞書とは、例えば機械翻訳の場合には翻訳対象言語に対する訳語の辞書である。18は学習辞書であって、曖昧性解消部16で曖昧性を解消された翻訳結果を学習し、後の曖昧性訳文に対して適切な訳文を見つけ出すときに用いられる。

【0021】前記通信部5は、入力部11から入力されるデータや記録部15に記録されているデータを自然言語処理端末3に送信したり、自然言語処理端末3からの送信データを受信して制御部13に送出する。この送信データは、曖昧性が存在する場所を示す情報を含むものである。

【0022】次に、自然言語処理端末3は、前述するように通信部6および自然言語処理エンジン7で構成され、そのうち自然言語処理エンジン7は、制御部21、

記録部22、翻訳部23および翻訳用辞書24によって構成されている。

【0023】通信部6は、通信部5から送信されてくる翻訳要求文やコマンドを受信したり、翻訳を終えたデータを送信したりする役目をもっている。翻訳処理によって曖昧性が生じたとき、その曖昧性を含めた翻訳結果を処理要求端末1に送信する。

【0024】自然言語処理エンジン7の中の制御部21は、記録部22や翻訳部23などと接続され、後記する図4の処理制御やデータの受け渡しを行う機能をもっている。記録部22は、通信部6で受信されたデータや翻訳部23で翻訳されたデータを一時的に保存する。翻訳部23は、翻訳用辞書24を参照し、通信部6で受信されたデータや記録部22に記録されたデータを翻訳する機能をもっている。

【0025】次に、以上のように構成された装置の動作について説明する。図3は処理要求端末1、つまりクライアント端末の処理手順の一例を説明する図である。

【0026】このクライアント端末側の制御部13は、入力部11から入力データがあったとき、当該入力データが翻訳用原文であるかどうか、例えば英日翻訳の場合には半角の有無等から判断する(ST1)。翻訳用原文であった場合、そのデータを記録部15に記録する(ST2)。ステップST1において入力データが翻訳用原文でないと判断された場合、クライアント端末に対するコマンドであると判断し、それが翻訳開始要求であるかどうかを判断する(ST3)。翻訳開始要求でなかった場合には、入力されたコマンドに応じた処理を行う(ST4)。翻訳開始要求であった場合には、記録部15に記録されているデータを自然言語処理端末3、つまりサーバ端末に送信する(ST5)。

【0027】ここで、クライアント端末側の制御部13は、翻訳用原文がサーバ端末で翻訳され、その翻訳結果が送信されてくるまで、データ待ちの状態となる(ST6)。サーバ端末の翻訳処理は後記する(図4参照)。なお、データ待ち状態の間、クライアント端末側は、他のコマンドや翻訳要求文の入力待ちをしていてもよい。

【0028】サーバ端末で翻訳が終了し、翻訳結果データが転送されてくると、それが曖昧性を含んだデータであるかどうかを判断し(ST7)、曖昧性を含んだデータである場合は、曖昧性解消部16が曖昧部分に対して曖昧性解消用辞書17に記載される辞書を参照し、曖昧部分に対する他候補を生成する(ST8)。つまり、辞書中に記載される同じ見出し語に対する他訳語を列挙し、表示部12に表示することにより、オペレータに翻訳結果を提示し(ST9)、翻訳結果に曖昧性が含まれているかどうかを判断する(ST10)。曖昧性が含まれている場合、オペレータからの入力待ち状態となる(ST11)。オペレータが表示されている候補の中から1つを選択すると、その候補と対応する曖昧部分の文字列とが

学習辞書18に保存する(ST12)。そして、選択された候補によって翻訳結果中の曖昧部分が置き換えられる(ST13)。

【0029】次に、図4にてサーバ端末の動作について説明する。サーバ端末側の制御部21は、クライアント端末からの翻訳用データを受信するために待機している(ST21)。翻訳用原文データを受信すると、そのデータを翻訳部23に転送する(ST22)。ここで、翻訳部23は、翻訳用辞書24に記載される辞書を参照しながら翻訳を行う。翻訳部23において翻訳処理を終了すると、翻訳部23から翻訳結果を受信する(ST23)。この翻訳結果の中に曖昧性を含んだデータであるかどうかを判断する(ST24)。この曖昧性の判断は、例えば1つの見出し語に対して複数の訳語がある場合に曖昧性ありと判断し、訳文中のその訳語に対して例えばマーキングなどの識別マークを施す(ST25)。そして、翻訳用データに対して全ての翻訳を行ったとき、その翻訳結果をクライアント端末に転送する(ST26)。

【0030】また、訳文野中に曖昧性があったとき、その訳語に対して例えばマーキングなどの識別マークでなく、その個所例えば何頁何行目とか、その他の方法で識別できるようにする。

【0031】さらに、本発明装置の具体的動作について図5ないし図8を参照して説明する。クライアント端末の制御部13は、翻訳用原文が入力されると、記録部15に記録する一方、図5に示すようなメモリ表示エリアの原文表示部Pに翻訳用の原文を格納し、表示部12に表示する。この原文表示部Pは、翻訳要求文の入力や編集を行うための画面として用いられる。画面下枠は、訳文表示部Qであって、翻訳結果やそれに付随する情報などを表示するために用いられる。

【0032】図6は、クライアント端末側表示部12の原文表示部P側に翻訳用原文、翻訳表示部Q側に翻訳結果がそれぞれ表示された状態を示している。この訳文表示部Qに表示される翻訳結果は第1解を表している。この翻訳結果の訳文のうち、例えば「あなた」には他にも訳語が存在するので曖昧性がある。

【0033】そこで、図6では、サーバ端末側から翻訳結果が送られてきたとき、その翻訳結果の訳文の中の「それ」、「その」、「出来事」、「に因んで」、「名付けられた」等にマーキングが施されているので、曖昧性があると判断する。訳文中にマーキングがある場合、クライアント端末側の曖昧性解消部16は、曖昧性解消用辞書を用いて他候補の生成を行うことになる。例えば画面上の「に因んで」を選択すると、他の訳語が1つ以上表示されるので、その訳語の内の1つを選択することにより、第1解に提示された語句と置換する。

【0034】図7は「に因んで」なる訳語に対する他候補の例であって、「にちなんで」、「の後に」、「のあ

とに」、「の後で」、「のあとで」が挙げられている。なお、処理するデータ量が少なければ少ないほど、曖昧性解消処理時間の短縮化を図ることができる。データ量を少なくする方法として、マーキングや曖昧性解消処理の対象となる曖昧性の種類をユーザカスタマイズ項目で設定すれば可能となる。例えば曖昧性解消処理の対象として、「名詞」のみと設定すれば、図6の訳文の場合には、「出来事」だけがマーキングされることになる。

【0035】また、曖昧性解消用データに対しても、表示順番や表示する種類を設定できるようにすることもできる。例えば曖昧性解消用データである訳語の他候補の表示順番を「漢字優先」と設定すれば、曖昧性解消部16は、当該設定内容に基づいて図7のプルダウンメニューの項目として、「に因んで」、「の後に」、「の後で」、「にちなんで」、…という順序で表示することができる。また、「イディオムのみ」と設定すれば、その設定内容に従って、「に因んで」以外に「にちなんで」しか表示されなくなる。

【0036】図8は図7の状況が発生したときにサーバから送信されたデータの一例を示す図である。この図に示すように、訳語一語が一行として表現されている。各行の第1カラムは曖昧性の有無を示すマーキング部分である。このマーキング部分は、ユーザカスタマイズ項目などを参照しながら付与される。第2カラムは、第一解の訳語である。この訳語の各々に対して、ユーザカスタマイズ項目の設定を参照しながら辞書引きが行われ、訳語の他候補が生成される。

【0037】従って、以上のような実施の形態によれば、サーバ端末は、クライアント端末から自然言語処理要求を受けたとき自然言語処理を行うが、この自然言語処理中に翻訳結果を一意に決定できない曖昧性の存在する訳文が生じたとき、曖昧性を解消せずにそのまま他の翻訳結果データと共に送信するので、サーバ端末の処理負担が非常に少なくなるばかりでなく、曖昧性を含んだ状態で直ちに翻訳結果を送信し、ネットワークの占有を解除するので、ネットワークの占有時間を大幅に低減化できる。このことは、サーバ端末に多数のクライアント端末が接続されていても、各クライアント端末の処理要求に十分に対応可能であり、システム全体のデータ処理の効率を高めることができる。一方、クライアント端末では、十分な時間をかけて曖昧性を解消する処理を行うことができる。

【0038】また、サーバ端末は、曖昧性の存在する訳文に、曖昧性の存在しない訳文とは区別できる識別データ付して送信するので、曖昧性を含む翻訳結果のデータをクライアント端末に確実に転送でき、また曖昧性を含む翻訳結果を受けたクライアント端末では識別データから容易に曖昧性の訳文を見つけ出して曖昧性の解消処理を行うことができる。また、クライアント端末は、サーバ端末から曖昧性を含んだ訳文を受けたとき、曖昧性解

消用辞書17を参照しつつ他候補を生成し、他候補の中から1つを選択したとき、その選択候補をサーバ端末に送信せずに自己の学習辞書18に保存するので、後の曖昧性を含む翻訳結果の解消時に学習結果を有効に利用でき、速やかに曖昧性を解消することができる。さらに、ユーザカスタマイズを設定し、整理した状態で曖昧性の解消を行えば、曖昧性を含む訳文のデータの曖昧性を速やかに解消でき、曖昧性解消処理データのデータ量を少くして曖昧性を解消できる。また、曖昧性解消用データを生成する際、ユーザカスタマイズ項目を設定することにより、ユーザが選択しやすいデータ表示とし、かつ、曖昧性解消用データのデータ量を低減化できる。

【0039】なお、本発明は、機械翻訳に限られるものではなく、例えば音声認識、スペルチェック、文章構成、検索等の処理にも同様に適用可能である。また、上記実施の形態は、英日翻訳の例を用いたが、その他の言語対に対しても同様である。また、日本語以外の言語に対する自然言語処理にも適用することができる。

【0040】

【発明の効果】以上説明したように本発明によれば、サーバ端末やネットワークの負荷を低減化でき、しかもサーバ・クライアントを効率よく利用しながら自然言語処理を行うことができる。

【図面の簡単な説明】

【図1】 本発明に係わる自然言語処理装置の概略構成図。

【図2】 図1に示す自然言語処理装置を具体化した一実施形態を示す構成図。

【図3】 クライアント端末の処理手順を説明するフローチャート。

【図4】 サーバ端末の処理手順を説明するフローチャート。

【図5】 クライアント端末における表示部の翻訳用原文の表示例を示す図。

【図6】 クライアント端末における表示部の翻訳用原文および曖昧性を含む翻訳結果を示す図。

【図7】 複数の曖昧性を含む訳文の中の1つの訳文について曖昧性解消用辞書から訳語を引き出して表示した図。

【図8】 サーバ端末から曖昧性を含む訳文を送信するときの一例図。

【符号の説明】

- 1…処理要求端末
- 2…ネットワーク
- 3…自然言語処理端末
- 4…データ処理制御部
- 5, 6…通信部
- 7…自然言語処理エンジン
- 13…制御部
- 16…曖昧性解消部

17…曖昧性解消用辞書

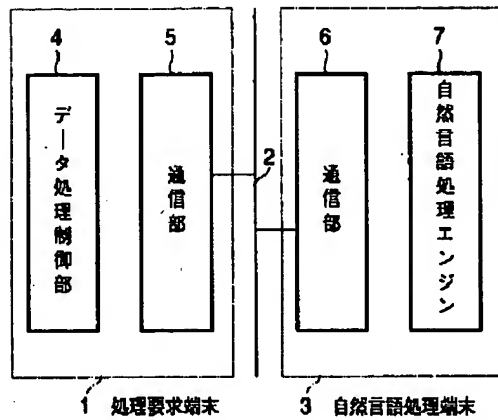
23…翻訳部

18…学習辞書

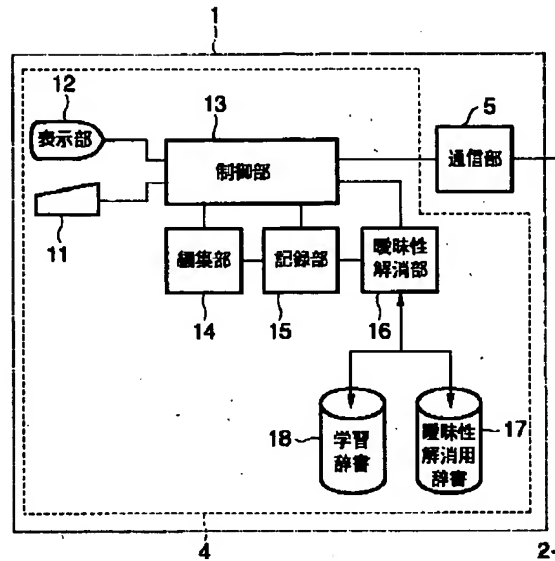
24…翻訳用辞書

21…制御部

【図1】



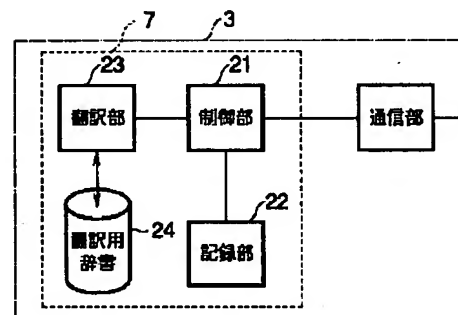
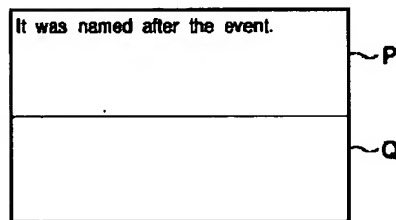
【図2】



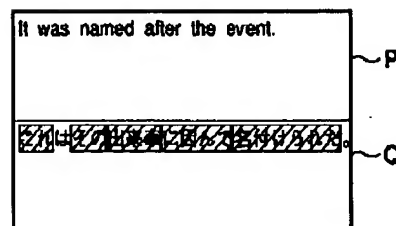
【図8】

●	それ
●	は
●	その
●	出来事
●	に因んで
●	名付けられた

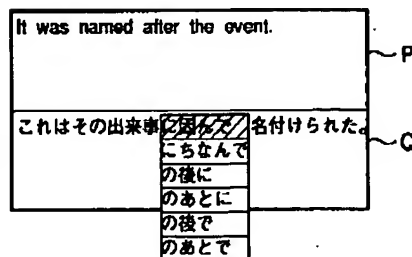
【図5】



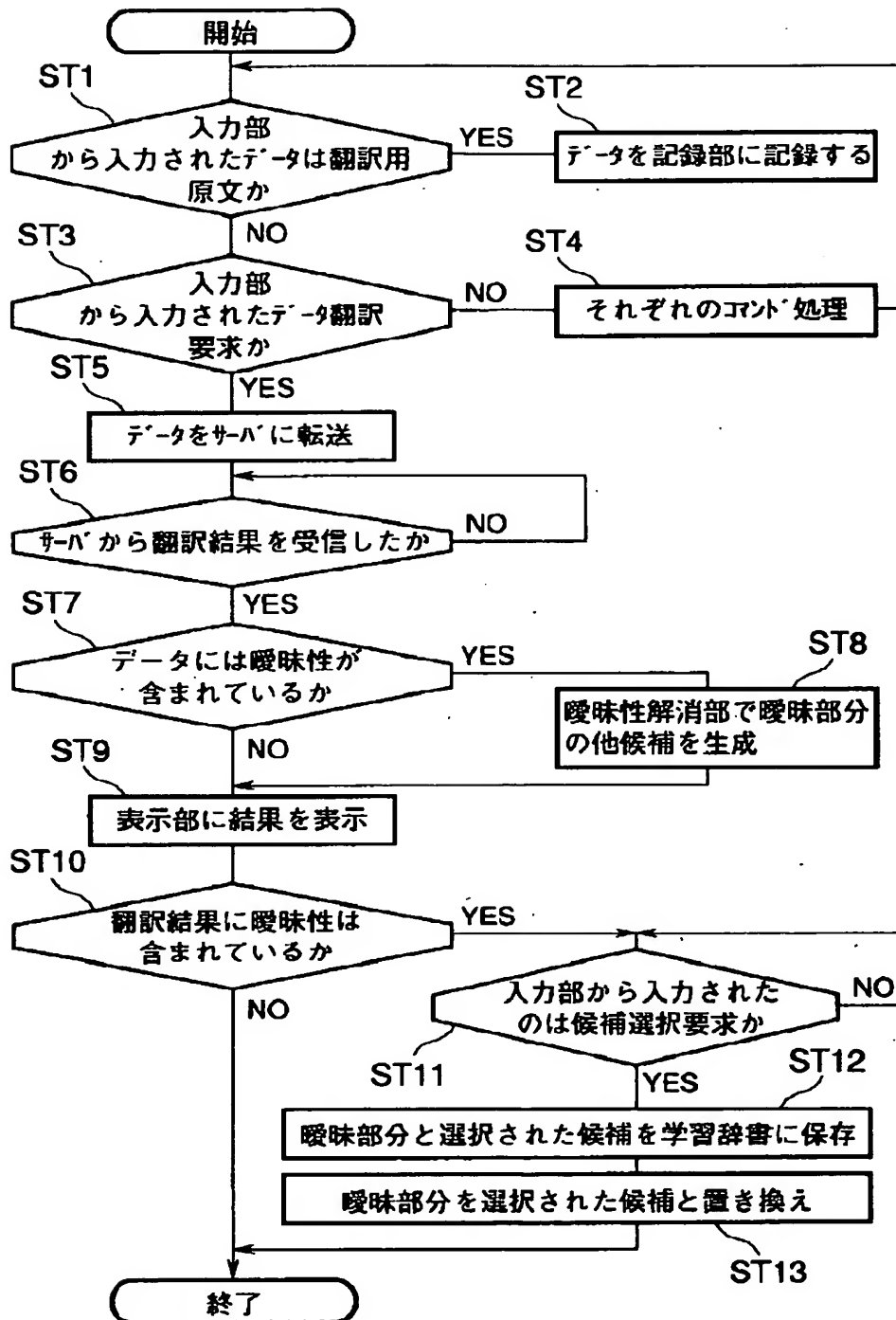
【図6】



【図7】



【図3】



【図4】

